

Génération de contraintes pour le clustering à partir d'une ontologie - Application à la classification d'images satellites

Hatim Chahdi^{*,**}, Nistor Grozavu^{**}, Isabelle Mougenot^{*}, Laure Berti-Equille^{*,***}, Younès Bennani^{**}

^{*}UMR U228 Espace-Dev, IRD - Université de Montpellier
Maison de la télédétection - 500 Rue J.F. Breton, 34093 Montpellier
prenom.nom@ird.fr,

^{**}UMR 7030 LIPN, CNRS - Université Paris 13
99, avenue Jean-Baptiste Clément, 93430 Villetaneuse
prenom.nom@lipn.univ-paris13.fr

^{***}Qatar Computing Research Institute, Hamad bin Khalifa University
Doha, Qatar

Résumé. L'utilisation des connaissances a priori peut fortement améliorer la classification non-supervisée. L'injection de ces connaissances sous forme de contraintes sur les données figure parmi les techniques les plus efficaces de la littérature. Cependant, la génération des contraintes est très coûteuse et demande l'intervention de l'expert ; la sémantique apportée par l'étiquetage de l'expert est aussi perdue dans ce type de techniques, seuls les contraintes sont retenues par le clustering. Dans cet article, nous proposons une nouvelle approche hybride exploitant le raisonnement à base d'ontologie pour générer automatiquement des contraintes permettant de guider et améliorer le clustering. L'utilisation d'une ontologie comme connaissance *a priori* a plusieurs avantages. Elle permet l'interprétation automatisée des connaissances, ajoute de la modularité dans la chaîne de traitement et améliore la qualité du clustering en prenant en compte la vision de l'utilisateur. Pour évaluer notre approche, nous l'avons appliquée à la classification d'images satellites et les résultats obtenus démontrent des améliorations notables à la fois au niveau de la qualité du clustering et au niveau de l'étiquetage sémantique des clusters sans intervention de l'expert.

1 Introduction

Ces dernières années, une grande quantité d'images satellites a été rendue disponible par les gouvernements et différents acteurs publics. L'analyse de ces images peut apporter des réponses tangibles à des problématiques d'ordre environnemental et sociétal. La segmentation est parmi les étapes les plus importantes du processus d'analyse d'images satellites et s'appuie essentiellement sur des algorithmes de clustering. Classiquement utilisé dans un cadre exploratoire et non supervisé, le clustering vise à partitionner de gros volumes de données non-étiquetées en un ensemble de groupes de données au regard de leurs similarités. Cependant, dans plusieurs cas, ce partitionnement est fortement lié à l'intérêt que porte l'utilisateur